

Maximizing graph probability under conditionally exponential models

STEFANO NASINI

Universitat Politècnica de Catalunya
stefano.nasini@upc.edu

Abstract

Designing reliable networks consists in finding topological structures, which are able to successfully carry out desired processes and operations. When this set of activities performed within a network are unknown and the only available information is a probabilistic model reflecting topological network features, highly probable networks are regarded as "reliable", in the sense of being consistent with those probabilistic model. In this paper we are studying the reliability maximization, based on the Exponential Random Graph Model (ERGM), whose statistical properties has been widely used to capture complex topological feature of real-world networks. Under such models the probability of a network is maximized when specified structural properties appear in the network. However, the search of maximally reliable (highly probable) networks might result in difficult combinatorial optimization problems and an important goal of this work is to translate them into solvable systems of linear constraints. Analytical and numerical results are provided, using exact optimization techniques and efficient computer implementation.

Key words: Network reliability; Random graphs; Combinatorial Optimization.

I. INTRODUCTION

Network reliability has been studied as the ability of a network to successfully carry out desired processes and operations. When those processes are only probabilistically known, the design of reliable networks might consist in the maximization of the probability of success of a given operation within a network or the minimization of its failure.

Sometimes the only available information is a probability distribution on the set of edges, so that the optimal decision of a global planner consists in the search of highly probable networks, under the specified probabilistic model.

The first and best known probabilistic model of network formation has been introduced by the seminal work of Erdős and Rainyi [3], who considered a fixed set of n nodes and an independent and equal probability $2d/(n(n-1))$ of observing edges among them. A related variant of this model considers networks chosen uniformly at random from the

collection of all graphs with n nodes and d edges. (For more details about network properties, see Bollobas [1], and Wasserman and Faust [10].)

This variant of the Erdős–Rainyi random graph might be regarded as a conditionally uniform random model and a straightforward generalization of it is to consider different conditioning information, other than the number of edges. A general approach to characterize families of networks with fixed conditioning information is based on generating integer solution of systems of linear constraints [2].

Let x_{ij} be the entries of the adjacency matrix (AM, from now on) of an undirected graph and \mathbf{x} the vector associated to its components sorted in lexicographic order. Castro and Nasini [2] considered the probability space (χ, p, \mathfrak{S}) , where χ is the set of all AMs of networks verifying a specified collection of linear constraints, that is to say $\chi = \{x_{ij} \in \{0,1\}, (i,j) \in \mathcal{H}^2 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ and $\mathcal{H}^2 = \{(i,j) : 1 \leq i \leq n-1, i < j \leq n\}$. The set \mathfrak{S} is a σ -algebra on χ and p a probabil-

ity measure. In the case of dealing with the set of networks with fixed number of edges, we have $A = [1 \dots 1]$.

The probability measure p might take different forms, capturing the processes and operations performed within the network. Exponential Random Graph Models (ERGM) has been widely used to capture the fact that the probability of a network should reflect its structural properties. Here we are considering Conditionally Exponential Random Graph Models (CERGM), which keeps the main probabilistic properties of the classical ERGM into a constrained sample space χ , as shown in the next section. Highly probable networks are regarded as "reliable" in the sense of being consistent with the real-world scenario captured by the specified probabilistic model, so that the problem of designing reliable networks is here translated into the one of maximizing graph probability under conditionally exponential models.

This analysis casts a mathematical bridge between probabilistic and optimization based models of network formation, which allow taking into account the emerging properties from the point of a global planner, who wish to allocate connections among nodes in such a way as to successfully carry out processes and operations performed within the networks.

The rest of this paper provides a mathematical programming based approach to deal with this problem, based on the linearization of complex combinatorial properties.

II. RELIABLE NETWORKS BY PROBABILITY MAXIMIZATION

Consider a collection of independent and identically distributed networks $\mathbf{x}_1, \dots, \mathbf{x}_N \sim p$. Let $S_j(\mathbf{x}_i)$, for $j = 1, \dots, s$, be a structural feature of network \mathbf{x}_i , for $i = 1, \dots, N$, and $\hat{\mu}_j = \sum_{i=1}^N S_j(\mathbf{x}_i) / N$ the empirical expectation of S_j , for $j = 1, \dots, s$. The ERGM arises as an answer to the question *can we recover p from $\hat{\mu}_1, \dots, \hat{\mu}_s$* ? A reasonable requirement a probability measure p must verify is that $\mathbb{E}_p[S_j(\mathbf{x})] = \int_{\chi} S_j(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \hat{\mu}_j$, for $f = 1 \dots s$, where

$\chi = \{x_{ij} \in \{0, 1\}, (i, j) \in \mathcal{H}^2 : A\mathbf{x} = \mathbf{b}\}$. (We are using the Lebesgue integral as a generalization of summation for continuous spaces, that is to say, spaces of valued networks, where χ is the set of valued AMs verifying a specified system linear constraints. In the cases of zero-one-networks considered in this paper the integral can be replaced by a summation.)

To solve this problem, the maximum entropy principle leads to pick the distribution $\widehat{p}(\mathbf{x})$, which verifies $\mathbb{E}_p[S_j(\mathbf{x})] = \hat{\mu}_j$, while maximizing the entropy:

$$\begin{aligned} \max \quad & - \int_{\chi} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \\ \text{s. to} \quad & \int_{\chi} S_j(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \hat{\mu}_j, \quad f = 1, \dots, s \\ & \int_{\chi} p(\mathbf{x}) d\mathbf{x} = 1. \end{aligned}$$

The classical method for solving this class of problems is to apply the Lagrange multipliers to each of the constraints and maximize the augmented functional with respect to $p(\mathbf{x})$. As described by Jaynes [5], applying the Euler equation of calculus of variation, the functional form of a conditionally exponential random graph model might be obtained in few algebraic steps:

$$\widehat{p}(\mathbf{x}) = \begin{cases} \kappa \exp(\boldsymbol{\theta} \mathbf{S}(\mathbf{x})) & \text{if } \mathbf{x} \in \chi \\ 0 & \text{otherwise} \end{cases}$$

The boldface symbols $\boldsymbol{\theta}$ and $\mathbf{S}(\mathbf{x})$ denote the vectors $[\theta_1, \dots, \theta_s]^T$ and $[S_1(\mathbf{x}), \dots, S_s(\mathbf{x})]^T$ respectively. The vector parameter $\boldsymbol{\theta}_j \in \mathbb{R}^s$ controls the tendency of networks with parameters $S_j(\mathbf{x})$ to be observed in the data. The scalar quantity $\kappa = \int_{\mathbf{x} \in \chi} e^{\sum_{j=1}^s \theta_j S_j(\mathbf{x})} d\mathbf{x}$ is known as the partition function. Many papers from statistical physics call the quantity $\sum_{j=1}^s \theta_j S_j(\mathbf{x})$ the *graph Hamiltonian* and denote it as $H(\mathbf{x})$.

The just described CERGM represents a generalization of the classical ERGM, where the sample space χ can be arbitrarily defined. If χ is the set of all simple graphs, then $|\chi| = 2^{\binom{n}{2}}$. If χ is the set of all undirected graphs with fixed number of edges d , then $|\chi| = \binom{n(n-1)/2}{d}$.

Finding a reliable network under the specified CERGM consist in maximizing $\kappa \exp(\theta \mathbf{S}(\mathbf{x}))$, subject to $\mathbf{x} \in \chi$.

The highly combinatorial nature of these problems is analyzed in the next section, based on different mathematical programming formulations.

III. NETWORK PROPERTIES BY SYSTEMS OF LINEAR CONSTRAINTS

Consider the set χ of undirected networks with fixed number of edges d and an exponential model on the sample space χ , with graph Hamiltonian $H(\mathbf{x}) = \theta \sum_{i < j < k} x_{ij} x_{jk} x_{ik}$, representing the number of closed triangles in the network. Maximizing the logarithm of the probability inside the specified sample space leads to the non-linear binary problem

$$\begin{aligned} \max \quad & \sum_{i < j < k} x_{ij} x_{jk} x_{ik} \\ \text{s. to} \quad & \sum_{i < j} x_{ij} = d \\ & x_{ij} \in \{0, 1\} \quad (i, j) \in \mathcal{H}^2. \end{aligned} \quad (2)$$

Let w_{ijk} be the binary indicator of the (i, j, k) triad, which is equal to one if a closed triangle between i, j and k exists, and zero otherwise. The following system of linear constraints characterize the state of w_{ijk} :

$$\begin{aligned} 1 - z_{ijk} &\leq x_{ij} \leq y_{ijk} \\ 1 - z_{ijk} &\leq x_{jk} \leq y_{ijk} \\ 1 - z_{ijk} &\leq x_{ik} \leq y_{ijk} \\ y_{ijk} - z_{ijk} &\leq w_{ijk} \\ w_{ijk} + z_{ijk} &\leq 1 \\ 3 - z_{ijk} &\geq x_{ij} + x_{jk} + x_{ik} \\ y_{ijk}, z_{ijk}, w_{ijk} &\in \{0, 1\} \end{aligned} \quad (3)$$

where y_{ijk} and z_{ijk} are auxiliary variables.

By introducing (3) in (2), for every (i, j, k) such that $1 \leq i \leq n-2$, $i < j \leq n-1$, $j < k \leq n$, and replacing the objective function $\sum_{i < j < k} x_{ij} x_{jk} x_{ik}$ with $\sum_{(i,j,k) \in \mathcal{H}^3} w_{ijk}$, the maximization of the graph probability under the specified model becomes a linear program in binary variables, which can be solved up to optimality by standard integer programming technics [8, 9].

Note that for particularly small d (high sparsity) an optimal network consist of a fully connected subgraph and several disconnected nodes. One possibility to overcome this drawback is either to include more network properties in the graph Hamiltonian or to redefine the sample space χ in such a way that the unwanted trivial solutions are discarded, for example by forcing *network connectivity*.

A straightforward way to algebraically force connectivity is to require the existence of a flow circulating within the network, from one node to all the others. Thus, we make use of an artificial flow of $n-1$ units, departing from one node $h \in \mathcal{V}$ and arriving to each of the $n-1$ remaining nodes. The existence of such flow is a sufficient and necessary condition for the network to be connected:

$$\begin{aligned} \sum_{j=1}^n f_{hj} - \sum_{j=1}^n f_{jh} &= n-1 \\ \sum_{j=1}^n f_{kj} - \sum_{j=1}^n f_{jk} &= -1 \quad k \neq h \\ f_{ij} + f_{ji} &\leq n x_{ij} \quad (i, j) \in \mathcal{H}^2 \\ f_{ij} f_{ji} &\geq 0 \end{aligned} \quad (4)$$

Based on the same network flow intuition the *average path length*, which is the average number of steps along the shortest paths for all pairs of nodes $-\frac{1}{n(n-1)} \sum_{i \neq j} \delta_{ij}$, where $\delta(i, j)$ is the shortest path between i and j – can be linearly modeled and introduced in the graph Hamiltonian.

To measure the minimum amount of circulating flow needed to deliver one unit of flow from one node to another, n types of commodity-flows are defined (one per each node) along with $n-1$ units of each type of flow departing from the corresponding node and reaching all the remaining nodes:

$$\begin{aligned} \sum_{k=1}^n f_{ij}^k + f_{ji}^k &\leq n^2 x_{ij}, \quad (i, j) \in \mathcal{H}^2 \\ \sum_{j=1}^n f_{hj}^h - \sum_{j=1}^n f_{jh}^h &= n-1, \quad h \in \mathcal{V} \\ \sum_{j=1}^n f_{kj}^h - \sum_{j=1}^n f_{jk}^h &= -1, \quad k \neq h \\ f_{ij} f_{ji} &\geq 0, \quad (i, j) \in \mathcal{H}^2 \end{aligned} \quad (5)$$

The total flow-based distance between nodes is then defined as $S(\mathbf{x}) = \min_f \sum_{(i,j) \in \mathcal{H}^2} \sum_{k=1}^n f_{ij}^k + f_{ji}^k$, subject to (5).

Another commonly used network parameter is the correlation between nodes properties or the physical distance between nodes. Consider a measure of distance between nodal properties δ . The total distance between nodes in the networks with respect to the space generated by the nodal properties is $S(\mathbf{x}) = \sum_{(i,j) \in \mathcal{H}^2} \delta_{ij} x_{ij}$.

Under the exponential family, network statistics must be additivity included in the graph Hamiltonian, as resulted from the maximum entropy principle: $\kappa \exp(\boldsymbol{\theta} \mathbf{S}(\mathbf{x}))$. Nonetheless, it has been widely discussed (see Handcock [4]) that this probabilistic models suffer of high *degeneracy* and lack of *robustness* (or *stability*). Degeneracy entails that the model places a disproportionate probability mass on only on a few of the possible graph configurations in χ . The lack of robustness can be seen from two different point of view:

- small changes in the parametrization $\boldsymbol{\theta}$ give rise to drastic changes in the optimal networks;
- several alternative optimal solutions with drastically different topological structures exist.

To deal with the robustness of the optimal solution, consider the maximal graph probability $P^* = \max H(\mathbf{x})$, s. to $\mathbf{x} \in \chi$, and suppose there exist multiple $\mathbf{x} \in \chi$ with the same optimal value P^* . A robust solution of the problem of maximizing graph probability under conditionally exponential models is given by a second stage max-min formulation, which use the previously found optimal solution as a lower bound:

$$\begin{aligned} \max \quad & H \\ & H \leq \theta_j S_j(\mathbf{x}) \quad j = 1 \dots s \\ \text{s. to} \quad & \sum_{j=1}^s \theta_j S_j(\mathbf{x}) \geq \gamma P^* \\ & \mathbf{x} \in \chi, \end{aligned} \quad (6)$$

where $\gamma \in [0, 1]$ is a tuning parameter to allow for sub-optimality. When $\gamma = 1$, the optimal

solution of (6) is a network which maximizes the conditionally exponential probability, while ensuring max-min decision criterion (as a robustness criterion) among the alternative optimal networks. When $\gamma = 0$, the optimal solution of (6) is a network which maximizes the graph probability based on a *non-linear-graph-Hamiltonian* $\kappa \exp(\min\{\theta_i S(\mathbf{x}) : i = 1 \dots s\})$, which doesn't belong to the exponential family.

The next section will numerically study the problems with the form of (6), with $\gamma = 0$.

IV. MAX-MIN FORMULATION AND NUMERICAL RESULTS

Let $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$ be the number of non-connected pairs of nodes $(n(n-1)/2 - \sum_{(i,j) \in \mathcal{H}^2} x_{ij})$ and number of triangles (characterized by (3)) of a network \mathbf{x} . We wish to maximize the graph probability within the sample space of all connected networks (characterized by (4)) and under a non-linear graph Hamiltonian with the form of $H(\mathbf{x}) = \min\{\theta_i S(\mathbf{x}) : i = 1 \dots s\}$. The resulting max-min formulation is the following:

$$\begin{aligned} \max \quad & H \\ \text{s. to} \quad & H \leq \alpha S_1(\mathbf{x}) \\ & H \leq (1 - \alpha) S_2(\mathbf{x}) \\ & \mathbf{x} \in \chi, \end{aligned} \quad (7)$$

where $\alpha = \frac{\theta_1}{\theta_1 + \theta_2}$ - a rescaling of the model parameters -.

The direct computation of the optimal solution of problem (7) by brunch and bound algorithm is time consuming when $n \geq 50$ and a strong lower bound of the optimal solution can be used to reduce the number of brunch and bound nodes.

Proposition 1. *The number of closed triangles in the optimal solution of problem (7) is bounded from below by $h = \min\left\{(n-1), \alpha \frac{(n-2)(n-1)}{2}\right\}$ and the number of edges is bounded from below by $(n-1) + h$.*

Proof. Since χ is the set of all connected graphs, we must have $\sum_{(i,j) \in \mathcal{H}^2} x_{ij} \geq n-1$. Every connected graph with $n-1$ edges - a tree - is associated to $H = 0$ in (7), as no closed triangle

exist in it. Consider a star as a feasible solution of (7) with $H = 0$ and add h edges between pairs of nodes at distance two. Since we have $n - 1$ pairs of nodes at distance two in a star, $h \leq n - 1$ additional edges might be included, creating h new triangle and resulting in an increase of the value of the objective function up to $\min \left\{ (1 - \alpha)h, \alpha \left(\frac{n(n-1)}{2} - (n-1) - h \right) \right\}$. Thus, we can keep increasing h as long as $(1 - \alpha)h \leq \alpha \left(\frac{n(n-1)}{2} - (n-1) - h \right)$, up to $n - 1$, so that the number of closed triangles and the number of edges in the optimal solution will be bounded from below by $h = \min \left\{ (n-1), \alpha \frac{(n-2)(n-1)}{2} \right\}$ and $(n-1) + h$, respectively. \square

Stronger valid inequalities can be numerically found by heuristic methods, such as local search, tabu search and ant colony¹.

Three instances of problem (7) are solved for $n = 60$ and $\alpha = 0.7, 0.5, 0.3$, as shown in figures 1, 2 and 3 respectively.

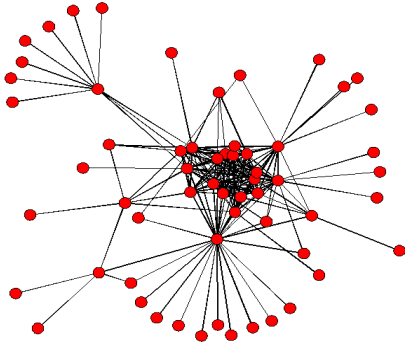


Figure 1: Optimal network, for $n = 60$ and $\alpha = 0.7$.

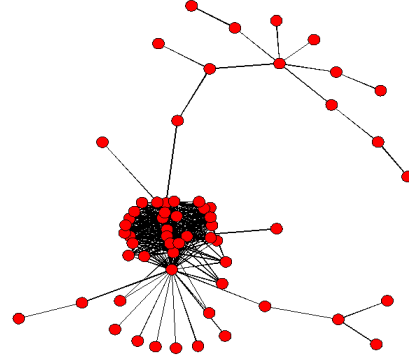


Figure 2: Optimal network, for $n = 60$ and $\alpha = 0.5$.

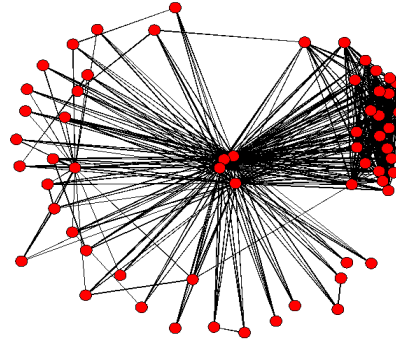


Figure 3: Optimal network, for $n = 60$ and $\alpha = 0.3$.

Table 1 reports the density, clustering coefficient and average path lengths of the three networks shown in figure 1, 2 and 3.

Dataset	Density	CC	APL
$\alpha = 0.7$	0.11921	0.61378	2.49265
$\alpha = 0.5$	0.17966	0.66311	3.17344
$\alpha = 0.3$	0.28192	0.63821	1.71808

Table 1: Resulting topological properties.

¹We implemented a first-improve local search which adds and remove one edge in each iteration up to converge to a local optimum, in which no single edge can be added or removed with an improve of the objective function.

Note that, although the average path length was not maximized, a really small degree of separation between nodes emerges, as shown in Table 1.

To obtain a deeper understanding about the emergence of a small degree of separation between nodes, consider an alternative model, where two measures of nodal distance are introduced:

- the physical distance $S_1(\mathbf{x}) = \sum_{(i,j) \in \mathcal{H}^2} \delta_{ij} x_{ij}$, where δ_{ij} is the distance between i and j ;
- the network distance $S_2(\mathbf{x})$, measured by the size of the total circulating flow to deliver one unit of flow from one node to another, as formulated in (5).

The max-min formulation in (6) with $\gamma = 0$ is considered² and three instances are solved for $n = 60$ and $\alpha = 0.7, 0.5, 0.3$, as shown in figures 4, 5 and 6 respectively³.

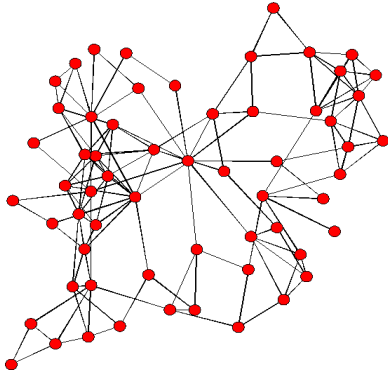


Figure 4: Optimal network, for $n = 60$ and $\alpha = 0.1$.

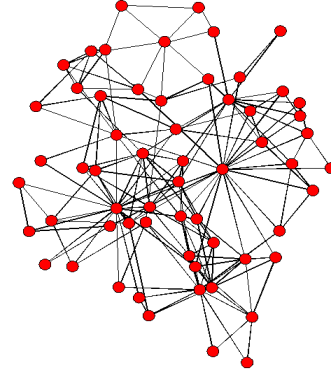


Figure 5: Optimal network, for $n = 60$ and $\alpha = 0.2$.

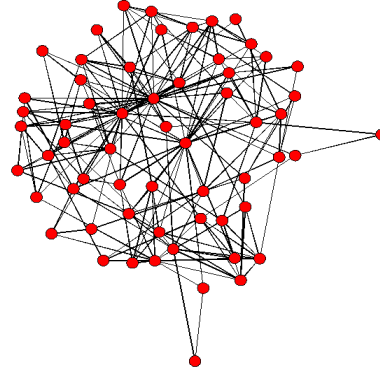


Figure 6: Optimal network, for $n = 60$ and $\alpha = 0.3$.

Although we only require to minimize the maximum between circulating total flows and total physical distance between connected nodes, the resulting networks shown in figure 4, 5 and 6 exhibit a remarkably high transitivity (high clustering coefficient), as shown in Table 2.

²In the case of dealing with negative model parameters θ_1 and θ_2 associated to the two types of distances the max-min formulation in (6) can be replaced by a min-max formulation, where the signs of θ_1 and θ_2 are switched to positive and then rescaled as $\alpha = \frac{\theta_1}{\theta_1 + \theta_2}$.

³This problem allows a straightforward application of the Benders decomposition, which separately and iteratively solves the problem of finding the minimum cost network structure and the one of setting the minimum circulating flow within it. We solve it by the brunch and bound algorithm, providing a lower bound to the optimal solution.

Dataset	Density	CC	APL
$\alpha = 0.1$	0.07910	0.41945	3.24633
$\alpha = 0.2$	0.09944	0.27842	2.49380
$\alpha = 0.3$	0.28192	0.23901	2.16554

Table 2: Resulting topological properties.

This result seems to be particularly coherent with the one of Mathias and Gopa [6], who showed that the small-world topology arises as a consequence of a tradeoff between maximal connectivity and minimal wiring.

Consider a toy model of the brain. Let us assume that it consists of local processing units, connected by wires. What constraints act on this system? On the one hand, one would want the highest connectivity (shortest path length) between the local processing units, so that information can be exchanged as fast as possible. On the other hand, it is wasteful to wire everything to everything else. (See Mathias and Gopa [6], page 2.)

The model of Mathias and Gopa tries to merge the intuition of a spatial distribution of nodes with the optimization based criterion of edge formation. Our approach in this context is purely based on the optimization of graph probability under a specified random model which reproduce the pattern of high transitivity and high connectivity of small-world network structures.

V. CONCLUSION

This work provided mathematical programming based approaches to deal with the problem of maximizing graph probability under conditionally exponential models. The underlying idea was to obtain highly reliable networks, when the set of operation and processes performed within a network are unknown and the only available information is a probabilistic model reflecting topological network features.

We have seen that the search of maximally reliable (highly probable) networks might result in difficult combinatorial optimization problems and an important goal of this work

was to translate them into solvable systems of linear constraints.

The ability of characterizing the sample space and the graph Hamiltonian by a well defined systems of linear constraints represented a necessary (and sufficient, due to their solvability by integer programming technics) condition to generate maximally reliable networks under a conditionally exponential model, as shown in the described computational results.

REFERENCES

- [1] Bollobas, B., (1985). *Random Graphs*, Cambridge University Press.
- [2] Castro, J., Nasini, S., (2014), Mathematical programming based approaches for classes of random network problems *Submitted to European Journal of Operational Research*.
- [3] Erdos, P., Rainyi, A., (1959). On random graphs, *Publicationes Mathematicae* 6: 290-297.
- [4] Handcock M., (2003), Assessing degeneracy in statistical models of social networks, *working paper n. 39*
- [5] Jaynes, E.T., (1957), Information theory and statistical mechanics, *Phys. Rev*, 106, 620.
- [6] Mathias N., Gopa, V., (2001). Small worlds: how and why, *Physical Review E*, 63, 021117.
- [7] Newman, M.E.J., (2002). Random graphs as models of networks, *Working Papers 02-04-020, Santa Fe Institute*.
- [8] Schrijver, A., (1998). *Theory of linear and integer programming*, John Wiley and Sons.
- [9] Schrijver, A., (2003). *Combinatorial Optimization*. Springer-Verlag.
- [10] Wasserman, S., Faust, K., (1994), *Social Network Analysis*, Cambridge: University Press.